Understanding and Improving Energy Efficiency of Regional Mobility Systems Leveraging System-Level Data

Sean Qian, Ph.D.
Director, Mobility Data Analytics Center (MAC)
Associate Professor, Carnegie Mellon University
seanqian@cmu.edu

June 2, 2020

Project ID:ti099

This presentation does not contain any proprietary, confidential, or otherwise restricted information

Overview

Overall goal:

- Optimal decision making for improving energy efficiency of mobility systems using multi-source system-level data

Timeline:

- Start: Jan 1, 2019
- End: Dec 31, 2021
- 40% complete

Budget:

- Total: $1,359,481
- DOE share: $1,000,000
- Cost share: $359,481
- Budget Period 1: $325,105
- Expended: $294,424
- Budget Period 2: $332,765
- Expended as of Apr 2020: $58,791

Partners:

- **Lead**: Carnegie Mellon University (CMU)
- National Renewable Energy Laboratory (NREL)
- Southwestern Pennsylvania Commission (SPC)
- Delaware Valley Regional Planning Commission (DVRPC)
- Pittsburgh Region Clean Cities (PRCC)
- The City of Pittsburgh

Barriers addressed:

- Difficult to acquire all relevant system-level data sets in a regional network
- Difficult to train high-resolution large-scale network flow models with multi-source system-level data
- Lack of comprehensive large-scale network models that help derive policies/decisions based on vehicle classifications (e.g., truck, EV, age, etc.) and Mobility energy productivity (MEP)

Project Objectives

Objectives

- Review inexpensive, replicable and openly-accessible data from multi-modal mobility systems, develop a data guide
- Develop a data-driven system-level modeling framework enabled and validated by large-scale data
- Identify the sources of energy inefficiencies of mobility systems from vehicles, passengers and infrastructure. (BP 3)
- Quantify the benefits of system-level strategies to improve mobility/energy efficiency, in Philadelphia and Pittsburgh regions (BP 3)

VTO TI goals

- Economic growth: the ability to model general trips and trucks
- Affordability for business and consumers: the ability to model system-level strategies/policies
- Reliability/resiliency: accessibility, day-to-day travel reliability, system resiliency to infrastructure disruptions

Impacts

- Comprehensive data from multiple sources, inexpensive and replicable for regions
- Hi-resolution network simulation and demand behavior that match large-scale data
- Models by vehicle classification lead to a better understanding of energy use and emissions
- System-level decision making on vehicles, demand and/or infrastructure (vehicle electrification, ride-sharing, parking pricing/availability, infrastructure expansions)

Project Approach 1

- Budget Period 1: Task 1- Data engine and mobility model
  - Task 1.1 - Database scheme and user interface
  - Task 1.2 – Data collection and processing (infrastructure, vehicle, passenger)
  - Task 1.3 - Data guide for system modeling and energy efficiency
  - Task 1.4 - Multi-class network simulation model
  - Task 1.5 - Behavior model: choices of routes, parking and time
  - **Data, data guide and web app are shared online**

- Budget Period 2: Task 2 – Large-scale deployment of regional mobility models and management strategies
  - Task 2.1 - Data-driven model calibration framework
  - Task 2.2 - Model calibration and deployment in Pittsburgh region (pending)
  - Task 2.3 - Model calibration and deployment in Philadelphia region (pending)
  - **Models, algorithms and regional case studies are all open sourced and will be shared online**

Project Approach 2

- Overall contributions
  - Comprehensive data in regional networks: weather, traffic counts/speed by vehicle classification, parking, vehicle registration/inspection (VIN etc.), GIS, transportation infrastructure improvement projects, incidents, etc.
  - A high-resolution network model (1-10 seconds, 100-1000 feet, EV, heavy-duty trucks, standard passenger vehicles, vehicle ages, etc.) with reasonable computational efficiency
  - Use machine learning theories to train the network model with multi-source data
  - Enables assessment of general system-level policies/strategies
  - Case studies in Pittsburgh and Philadelphia regions
  - Different modeling focus from POLARIS and BEAM

Milestones 1

Budget Period 1: (all milestones completed)

Milestone	Type	Description
Database and User Interface design Completed	Technical	The database and corresponding user-interface design is completed. The database with synthetic or partial datasets is implemented and tested.
All datasets stored	Technical	All datasets are processed and stored.
Data Guide Completed	Technical	A comprehensive data guide for energy efficiency study is completed.
Network Simulation Model Design Completed	Technical	The design and implementation of the network simulation model is completed and the model is tested on synthetic networks.
Multi-Class Dynamic Network Simulation Model	Go/No Go	A multi-class dynamic network simulation model that encapsulates propagations of passenger/vehicular trips in the roadway-parking network and their associated energy use in high spatio-temporal resolutions is implemented and tested in the synthetic networks.

Milestones 2

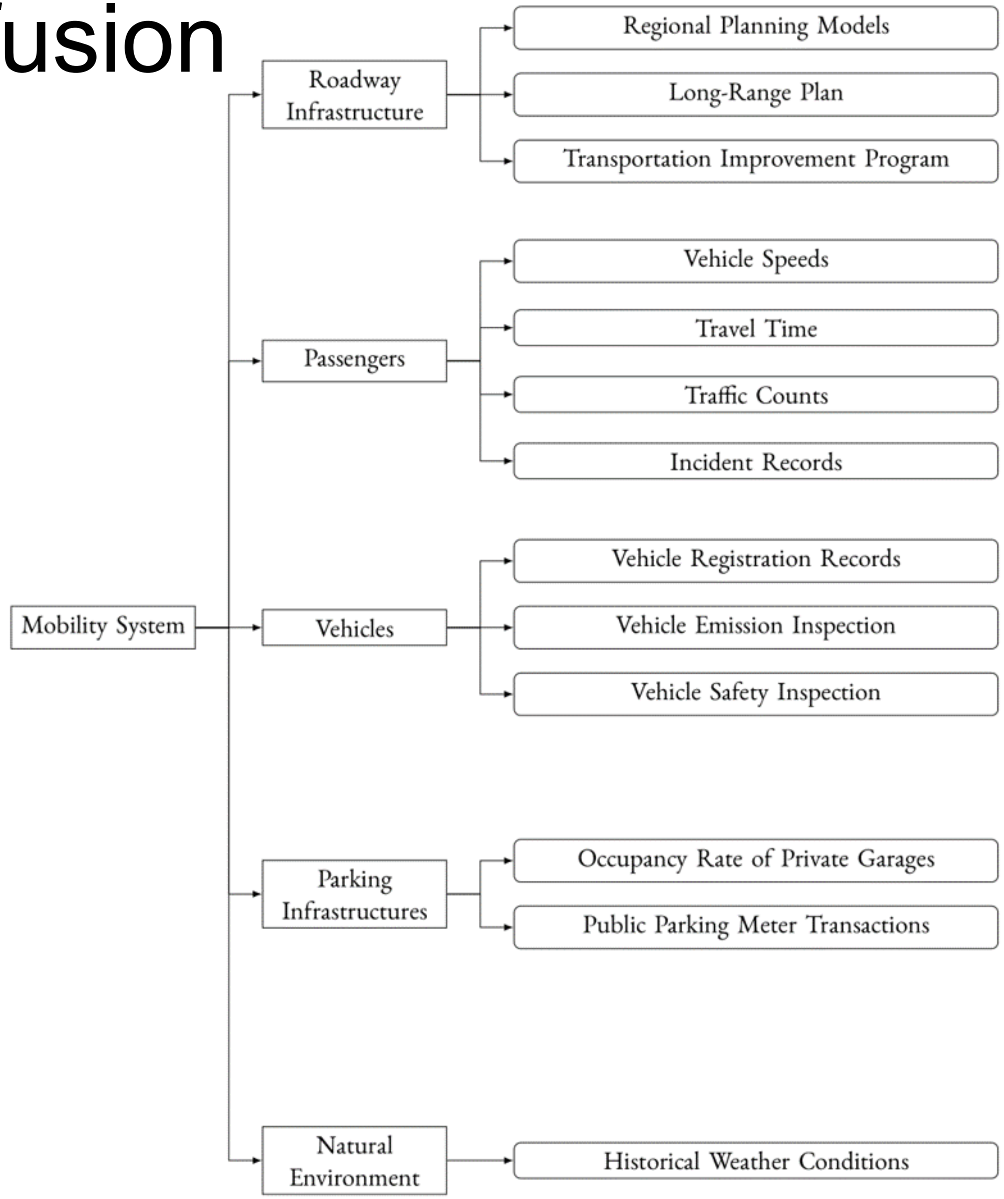
Budget Period 2:

Milestone	Type	Description
Model Calibration Framework Completed	Technical	The data-driven model calibration framework is designed, modeled and implemented. The framework is tested on small/large networks with synthesized data.
Dynamic Network Model Calibration Initiated	Technical	The calibration of the dynamic network model is completed. The R square score between the simulated (estimated) passenger/vehicular flow and the observed flow shall be no less than 0.6.
Management Strategies Implemented	Technical	The management strategies for energy efficiency in mobility systems are designed, modeled and implemented.
Go/No-Go # 3	Go/No Go	Calibration of the dynamic network model in demonstration networks is completed. The R square score between the simulated (estimated) passenger/vehicular flow and the observed flow shall be no less than 0.6.

Highlighted in blue are future tasks

Project Accomplishments: Data fusion

- Roadway infrastructures
  - Regional planning models
  - Transportation improvement plans
- Vehicles
  - State vehicle registration records
  - Vehicle safety and emission inspection results
- Passengers
  - Traffic speeds and travel times, by car and trucks
  - Vehicle counts, by car and trucks
  - Incident records
- Parking (Pittsburgh only)
  - On-street parking transactions
  - Private garage occupancy rates
- Weather



Project Accomplishments: Data Guide

- Brief summary of data
  - Information provided
  - How the data are collected
- How to obtain the data
  - How to get the raw data
  - How to store the data
- How to use the data
  - Spatial and temporal resolution of the data
  - Recommended preprocessing procedures
- Caveats and suggestions

3.1.1 INRIX TMC Based Probe Vehicle Speed Records

Summary INRIX[4] is a private company providing data for transportation analytics and planning. Unlike the traditional methods for collecting roadway data, INRIX speed data are from probe vehicles, making its spatial coverage and granularity outperform other sources of data. INRIX provides both historical vehicle speed records and the real time speed of a roadway segment. In this project, we will mainly use its historical data archive.

TMC (Traffic Message Channel) is a way to split roadway into segments. It is also used by other vehicle speed data providers like HERE, making it a good choice if we want to merge multiple sources of data. The INRIX TMC covers all highways and major arterials and are available from 2011.

How to Obtain Therefore multiple ways to access the INRIX TMC speed data, and we used the Regional Integrated Transportation Information System (RITIS)[5]. Here we list the steps to get the data:

1. Login to RITIS using a valid account;
2. Go to the RITIS Probe Data Analytics Suite Massive Data Downloader[6]. The interface is shown in Figure 8;
3. Select TMC segments from INRIX;
4. Select the road segments of interest on the map;
5. Specify the time range, temporal granularity, and features;
6. Submit the request. The data set can be found and downloaded from the "My History" page[7] once the request has been processed successfully;

Depending on the region and time range, the downloaded data files could be very large. So we recommend to download the raw data instead of the imputed data from INRIX.

How to Use The downloaded file will be a compressed archive and there are four files in it:

- `Contents.txt` shows a summary on the data downloaded;
- `Readings.csv` contains the actual data;
- `TMC_Identification.csv` is for looking up a segment via its TMC ID. It contains the name of the road and the longitude and latitude of both ends of the segment.

There are eight columns in the `Readings.csv` file, which are:

[4] http://www.inrix.com/
[5] https://www.ritis.org/
[6] https://pda.ritis.org/suite/download/
[7] https://pda.ritis.org/suite/my-history/

Figure 8: RITIS Probe Data Analytics Suite, Massive Data Downloader interface.

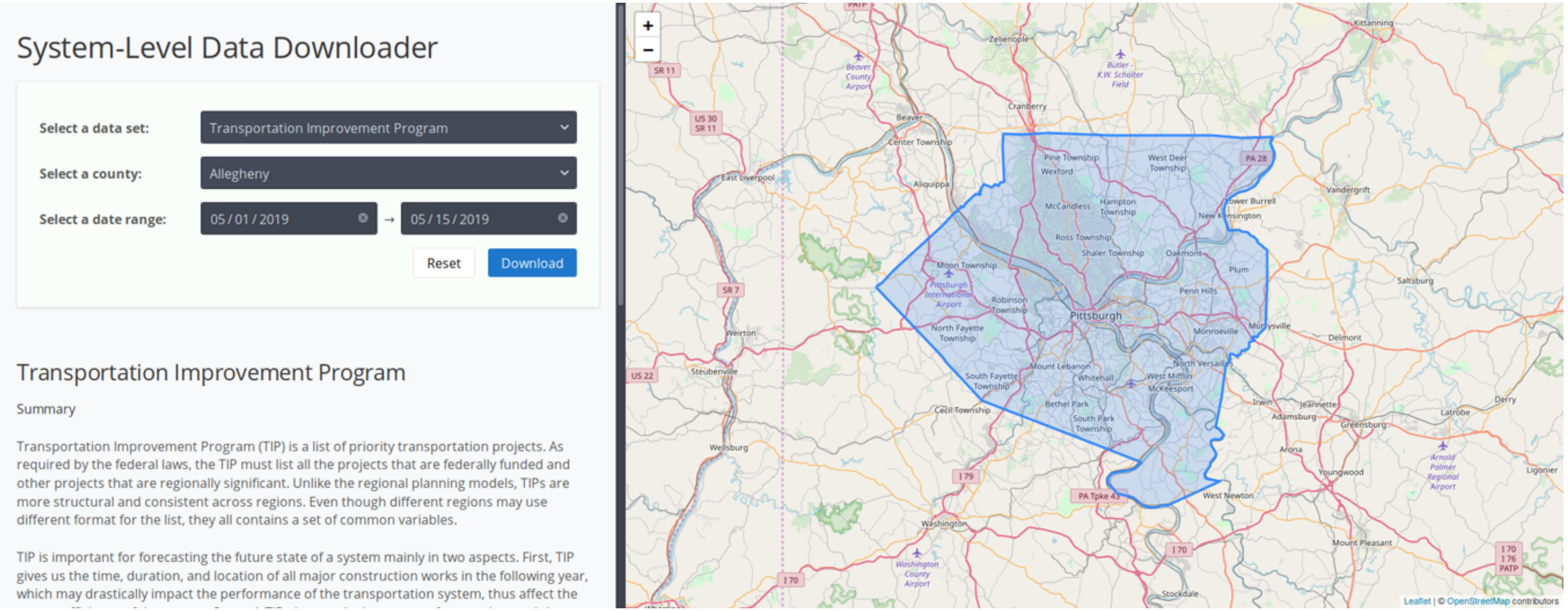
code: the ID of the roadway segment. The information of the segment can be found in the Identification.csv file;
urement_tstamp: the timestamp for the record;
d: the actual average speed in miles per hour within the time interval;
age_speed: the historical average speed of the segment at the same time of the day;
rence_speed: the typical vehicle speed under free flow conditions;
el_time_minutes: the travel time on the segment during the corresponding time interval;
ifence_score: the source of the record. 10 means the speed is the reference speed of that ent, 20 means it is the historical average speed, and 30 means it is calculated from the real-time e vehicle speed. It could also be a number between two categories, such as 25, and that means the record is from a mixed sources;

- `cvalue`: the probability (from 0 to 100) that the speed value represents the actual roadway conditions. It is only available when the record is from the real-time probe vehicle readings, i.e., the confidence score of the record is 30.

INRIX XD XD is another way for defining segments of a roadway. It applies a finer level of spatial granularity than TMC and has a larger coverage of roadways. It was launched in 2013 and is being actively developed. Usually there are four updates per year on the XD maps. The XD based INRIX data contain almost the same entries as the TMC based data–timestamp, real time speed, historical average speed, reference speed, and data quality.

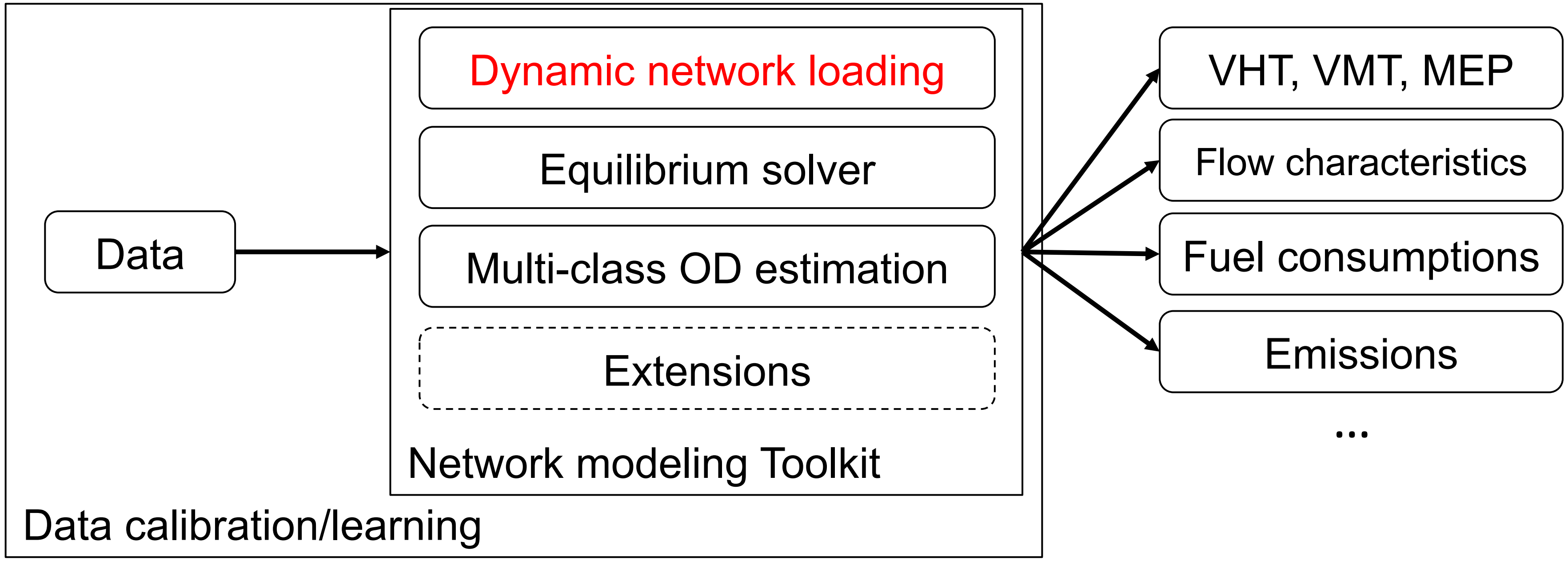
Similar to the TMC based data, the INRIX XD based data can be acquired from multiple sources as well. For

Project Accomplishments: Web app



Project Accomplishments: multi-class network models are completed 1

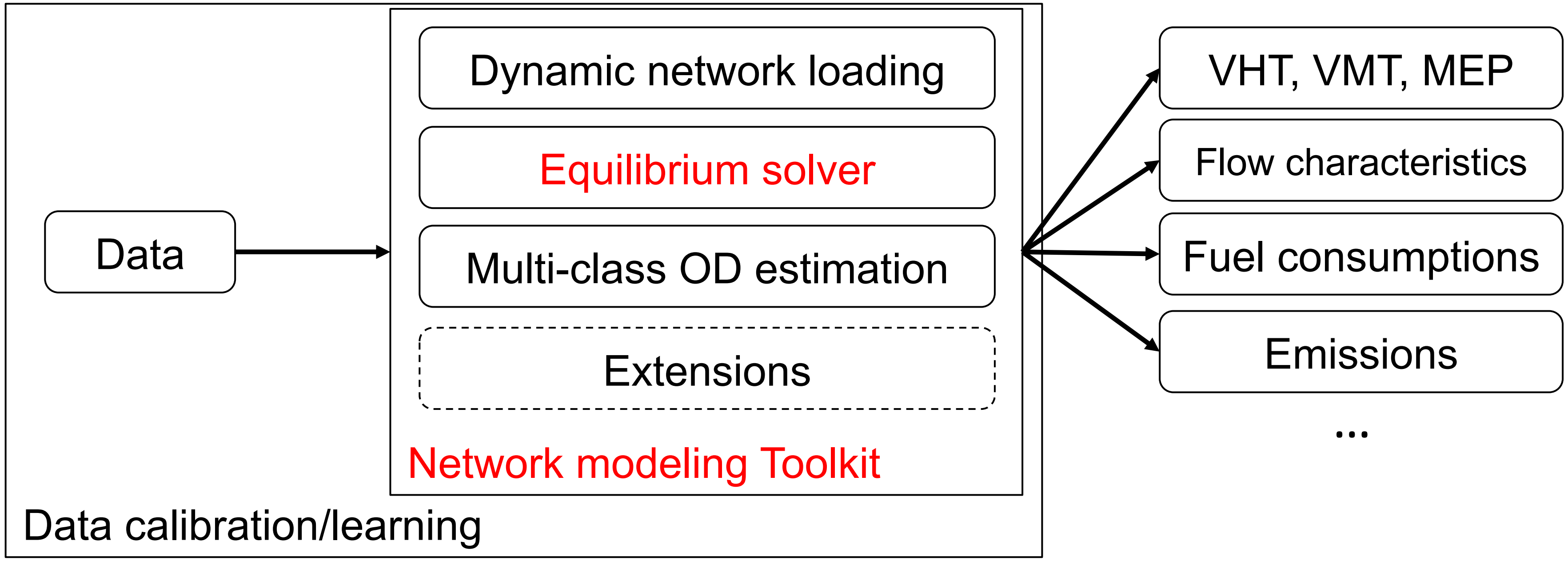
Tested in both Synthesized networks and real-world networks



- Network loading features
  - Cell transmission models, Link transmission models, Link queue models
  - Model car-truck interactions on roadways
  - Multi-class vehicles (trucks, diesel, EV, etc.)
  - Parking availability/cruising is explicitly modeled
  - **Project progress:** Efficient for large-scale networks (Pittsburgh, Philadelphia)
  - Use vehicle trajectories to estimate fuel use, emissions, MEP and other metrics

Project Accomplishments: multi-class network models are completed 2

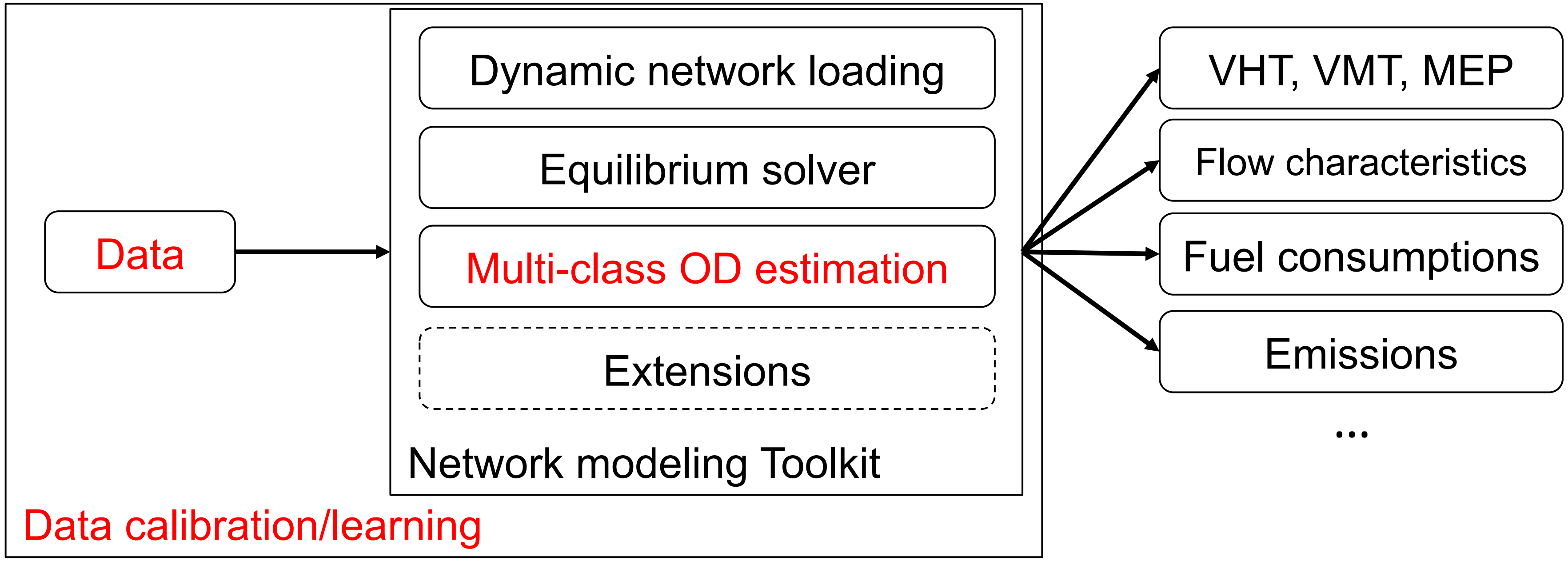
Tested in both Synthesized networks and real-world networks



- Behavioral modeling features
  - Each vehicular trip chooses his/her route, parking and time in such a way that all trips reach Equilibrium (or stochastic)
  - Mathematically formulated by a Variational Inequality problem and solved using optimization algorithms
  - **Project progress:** Need to improve the efficiency for large-scale networks (Pittsburgh, Philadelphia)

Project Accomplishments: multi-class network models are completed 3

Tested in both Synthesized networks and real-world networks



- Data learning features
  - Weather and incidents data: separate traffic patterns
  - Passenger, vehicle and infrastructure data: estimate multi-class origin-destination demand in a way to fit data in large-scale
  - Machine learning approach: a computational graph framework that incorporates general data input and works with large networks
  - **Project progress:** Efficient for large-scale networks (Pittsburgh, Philadelphia)
  - **Project progress:** R2 = 0.61 overall for Pittsburgh (100,000+ demand/vehicle spatio-temporal tagged data, and to be further improved)

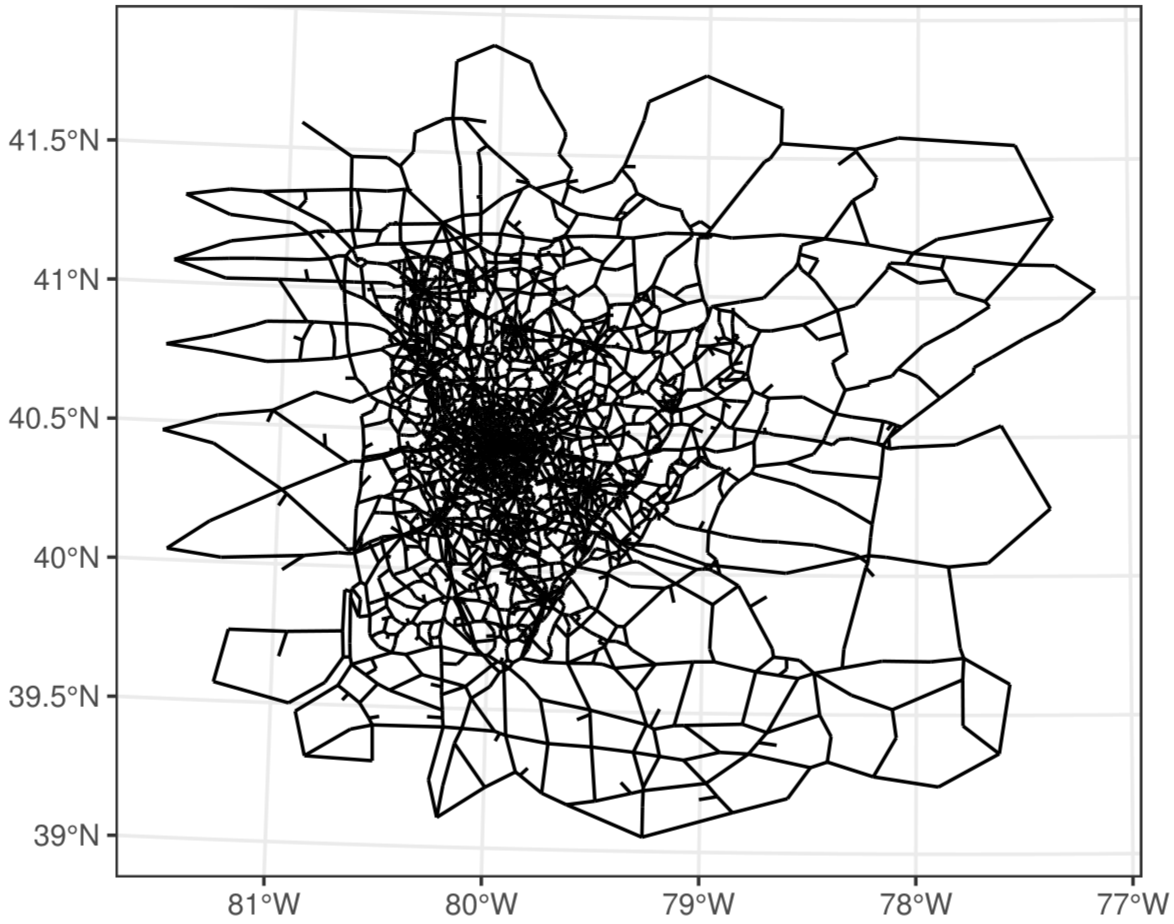
Project progress: an open-source tool

- MAC-POSTS
  - CMU Mobility data Analytics Center - Prediction, Optimization, and Simulation Toolkit for transportation Systems

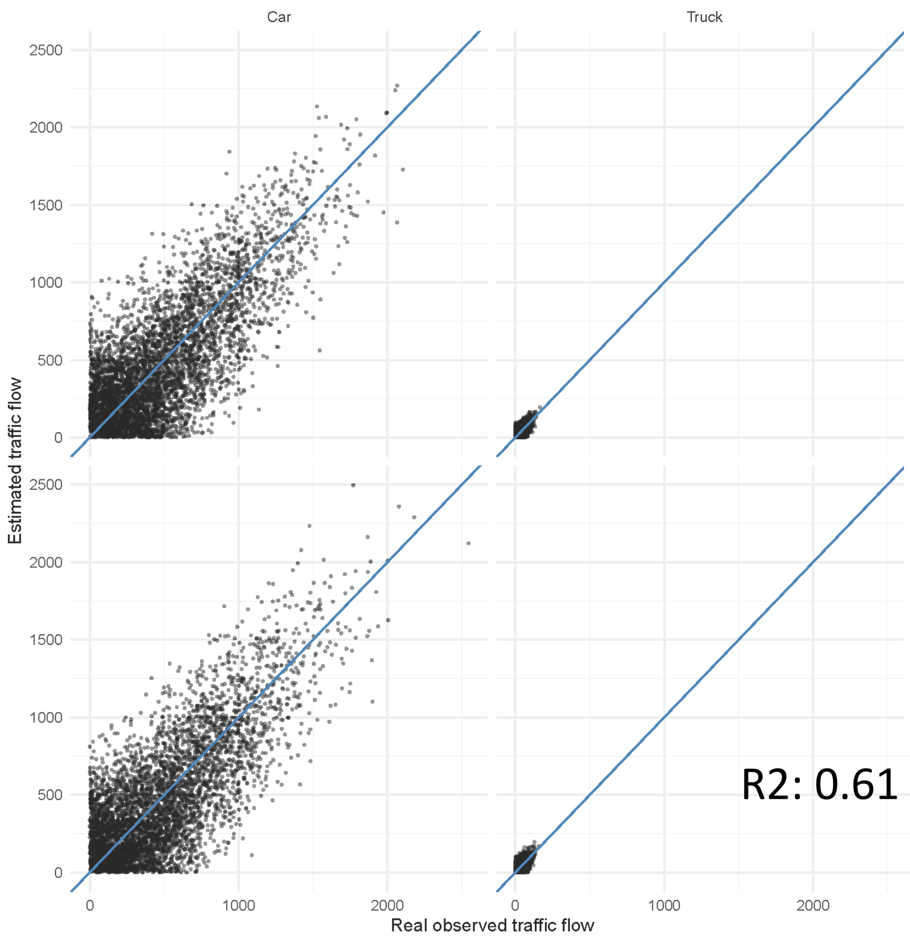
(It was initially supported by US DOT and NSF)

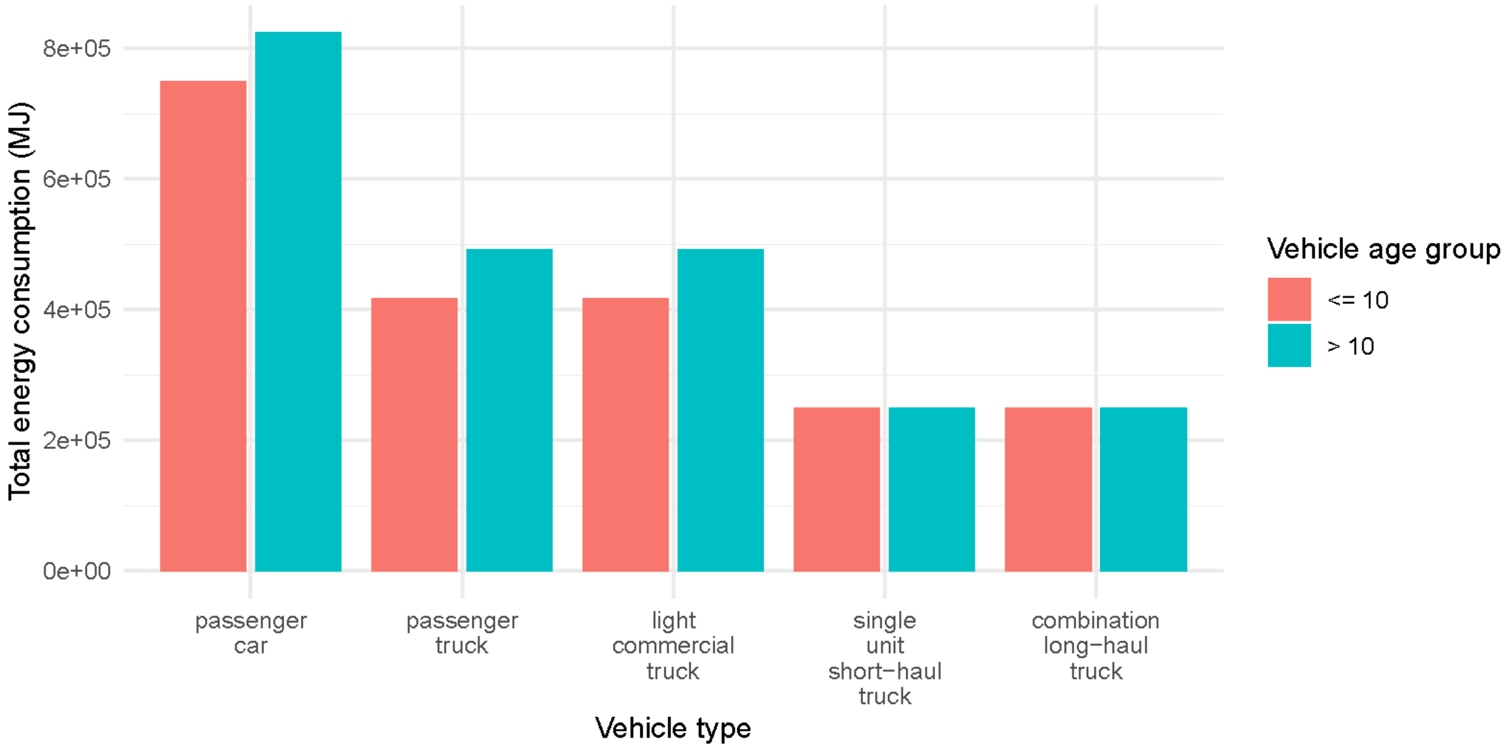
Project progress: Pittsburgh region case study

- Tested through synthetic networks with 99% R2
- Pittsburgh metro area
  - Weekday, 6am-noon
  - 5 seconds
  - 16,100 links, 6,297 nodes
  - 80,089 O-D pairs
  - 0.5 million trips
  - Data: 15-min car/truck counts, 5-min car/truck speeds, 5-min parking, vehicle registrations, roadways, weather, etc.



Project progress: Pittsburgh region case study





Preliminary, need to be improved and fine tuned

Collaboration and Coordination

- Biweekly meetings
- BOX fileshare
- Quarterly progress reports
- Other relevant projects leveraging the data/models developed through this DOE project

Sponsor: DOE VTO

Lead: Carnegie Mellon University (CMU)
- National Renewable Energy Laboratory (NREL)

CMU (co-advise two phd students):
Sean Qian: PI, network modeling, traffic engineering, multi-class modeling, parking analysis, data analytics, machine learning
Chris Hendrickson: data analytics, mobility policies
Jeremy Michalek: vehicle electrification
Costa Samaras: energy policies
Scott Matthews: life cycle analysis, data analytics
NREL: Josh Sperling and Stan Young: MEP implementation, parking, outreach

Partners:
- Southwestern Pennsylvania Commission (SPC)
- Delaware Valley Regional Planning Commission (DVRPC)
- Pittsburgh Region Clean Cities (PRCC)
- The City of Pittsburgh

Overall Impact

- Acquired, reviewed, fused and analyzed comprehensive data from multiple sources, inexpensive and replicable for regions.
- Demonstrated the possibility of replicating the work in other regionals using general data sets.
- Developed hi-resolution network simulation and demand behavior models that match large-scale data
- The models enables assessment of system-level strategies/policies with a better understanding of mobility systems and energy (in)efficiency
- Upcoming:
  - Fully demonstrate its modeling accuracy and effectiveness in Pittsburgh and Philadelphia regions
  - System-level decision making on vehicles, demand and/or infrastructure (vehicle electrification, ride-sharing, parking pricing/availability, infrastructure expansions)

Summary

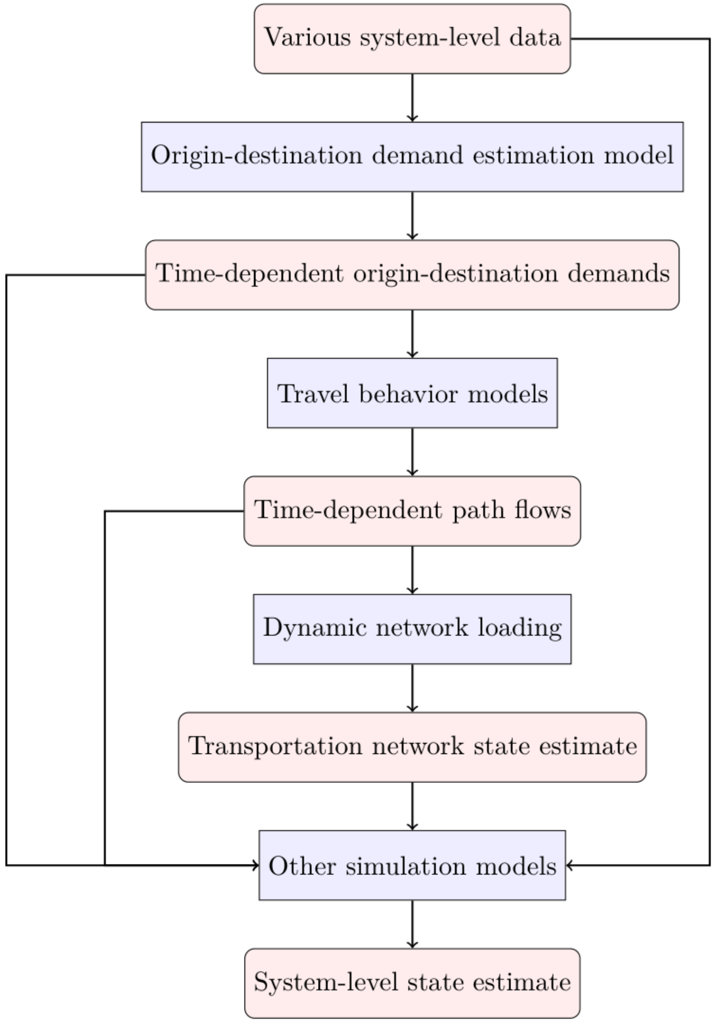
- System-level data sets in a regional network
  - Infrastructure, vehicles, demand, parking, weather
  - Mesoscopic simulation + system-level travel behavior
- Develop a machine learning approach to train high-resolution large-scale network flow models with multi-source system-level data
  - Multi-source data matching is the key, so we have the confidence on models for future prediction
- Assess/derive policies/decisions based on vehicle classifications (e.g., truck, EV, age, etc.) and Mobility energy productivity (MEP)
- Replicability: data base, user interface, data guide, general and efficient models, and general strategies/policies evaluation

TECHNICAL BACK-UP SLIDES

A computational graph approach to calibrate multi-class dynamic networks

- We proposed a theoretical formulation for estimating multi-class dynamic OD demand. The formulation is represented on a computational graph such that the multi-class dynamic origin-destination estimation (MCDODE) can be solved for large-scale networks with large-scale traffic data.
- The proposed MCDODE formulation can handle any form of traffic data, such as vehicle registration data, flow, speed or trip cost.
- We propose a novel forward-backward algorithm to solve for the MCDODE formulation on the constructed computational graph with simulation-based traffic assignment models. Due to the analogy between computational graphs and neural networks, many techniques used for deep learning can be used to solve MCDODE efficiently in large-scale networks.
- For example, we demonstrated that the advanced gradient-based methods and multiprocessing can significantly improve the efficiency of MCDODE in real-world networks.
- We use a tree-based cumulative curves to evaluate the gradient of multi-class dynamic OD demand in the forward-backward algorithm, and a Growing Tree algorithm is proposed to construct the tree-based cumulative curves during the multi-class network simulation.
- We examine the proposed MCDODE framework on a large-scale network (currently in Pittsburgh region) to demonstrate its effectiveness and computational efficiency of the solution algorithm. The results are promising.

Conceptual work flow for replicating this work in other regions



Using vehicle registration data in multi-class traffic network modeling

- PA has about 11.8 million registered vehicles
- Each vehicle has an associated zipcode (primary residence)
- About 40% of those VINs can be decoded.
- Retain vehicles with annual mileages greater than 4,200
- The network model is calibrated to match the total number of vehicles in each of the 10 vehicle classes for each zipcode

Policy/strategy scenarios

Create scenarios to estimate energy inefficiency (what sector is the most MEP-effective to improve energy efficiency). Some scenarios:

- Vehicle electrification
  - Replace current vehicular demand with electrical vehicles
- Ride-sharing
  - Reduce current vehicular demand by trip consolidation
- Parking pricing/availability
  - Build parking prices into individual choice model
  - Change parking inventory settings
- Infrastructure expansions
  - Change infrastructure (main roads) characteristics, such as the number of lanes, speed limit.

REVIEWER-ONLY SLIDES

Publications and Presentations

- Pengji Zhang, Sean Qian, (2019) “System-Level Data Guide for Modeling Regional Mobility Systems”, DOE technical report
- Wei Ma, Sean Qian, (2020) “Estimating multi-class dynamic origin-destination demand through a forward-backward algorithm on computational graphs”, under review, submitted to Transportation Research Part C
- Pengji Zhang, Sean Qian, (2020) “Network flow-based clustering and hypothesis tests”, in preparation, to be submitted to Transportation Research Part C
- Pengji Zhang, Sean Qian, (2020) “Network modeling with vehicle registration data”, in preparation, working paper (also presented in a CMU workshop)
- Pengji Zhang, Sean Qian, (2020) “Estimating marginal path cost in large-scale networks with multi-class traffic flow”, in preparation, working paper
- Chris Hoehne, Josh Sperling, Stan Young, Venu Garikapati, Sean Qian, (2020) “Parking as a lens to the urban soul: exploring associations of parking, mobility, and energy”, accepted and to be presented in the 27th Intelligent Transportation Systems World Congress
- CMU and NREL team have jointly presented those research work to DOE project managers, 2020 Transportation Research Board Annual Meeting, 2019 INFORMS Annual Meeting.

Critical Assumptions and Issues

- Assumptions
  – Choices of transportation modes, such as buses, bikes and ride-sharing are not considered. In other words, we assume travel demand on cars and trucks are fixed from day to day. This can be further relaxed under this modeling framework in future research.
  – Emissions and fuel use are estimated using MOVES Lite.
- Issues are mostly related to the network model to be further improved in BP 2
  – Network gridlock is a common issue in network modeling. We have ways to mitigate it, but cannot eliminate it.
  – The memory space and runtime of the models for large-scale networks need to be improved.
  – Though the goodness of fit for large-scale networks is reasonably good, we still hope to develop better algorithms to improve it.
  – Data loading to the network models is not automated, which may hinder the replicability of the data-driven model in other regions. This will be addressed in BP 3.